

用電量預測

純數據分析研究與實作

Victor

DATASET

訓練集：

	時間	溫度	變數1	壓力	風速	變數2	用電量
ID	datetime	temperature	var1	pressure	windspeed	var2	electricity_consumption
0	2013-07-01 00:00:00	-11.4	-17.1	1003.0	571.91	A	216.0
1	2013-07-01 01:00:00	-12.1	-19.3	996.0	575.04	A	210.0
2	2013-07-01 02:00:00	-12.9	-20.0	1000.0	578.435	A	225.0
3	2013-07-01 03:00:00	-11.4	-17.1	995.0	582.58	A	216.0
4	2013-07-01 04:00:00	-11.4	-19.3	1005.0	586.6	A	222.0
5	2013-07-01 05:00:00	-10.7	-19.3	1013.0	2.79	A	216.0

資料自2013年7月1日至2017年6月23日每小時記錄，共26497筆

特徵輸入值：時間、溫度、壓力、風速、var1、var2

預測目標：用電量

使用演算法 預測

- Decision tree - 決策樹
- XGboost – 極限梯度提升
- Linear regression(Lasso, Ridge, Elasticnet) - 線性回歸
- KNN – K 近鄰演算法
- SVR - 支援向量回歸
- MLP - 多層感知器

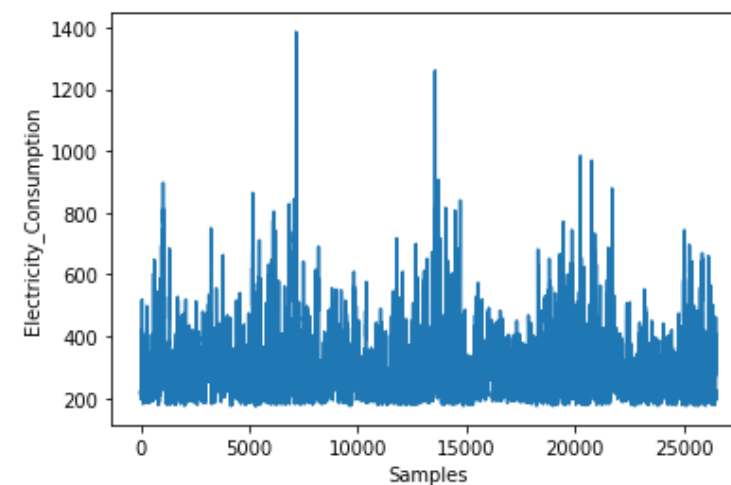
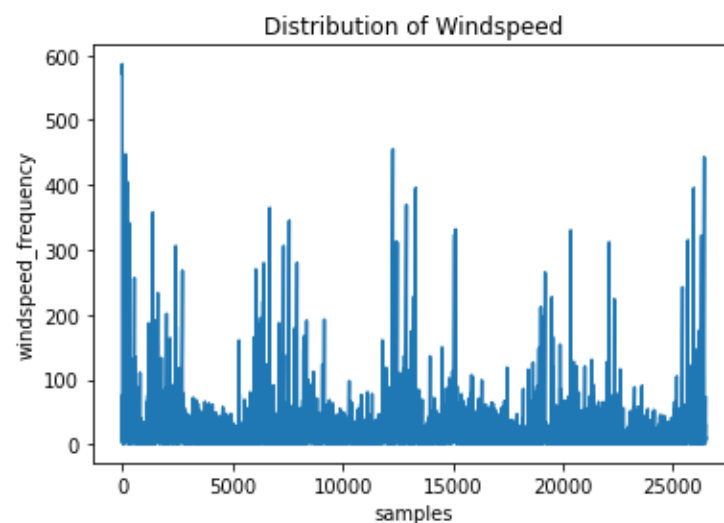
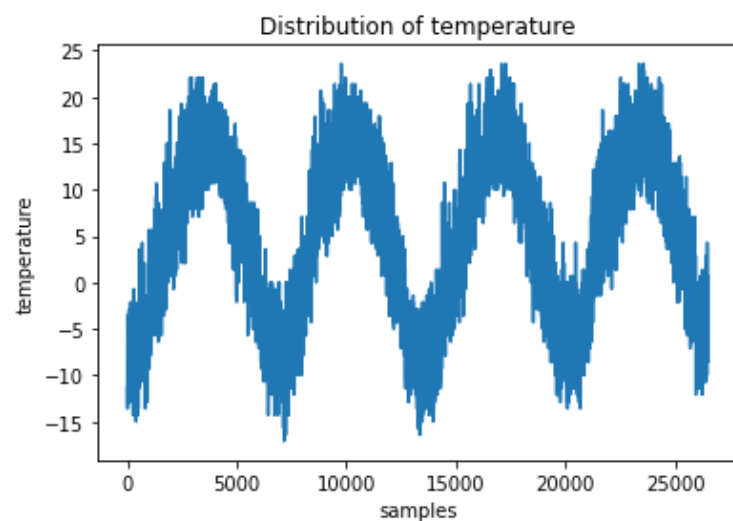
經訓練和驗證後效果最好的演算法為 XGboost和MLP ， MAE分別為54 、 55 。

XGboost

使用XGboost回歸演算法
預測用電量，並看出各特
徵的重要性，可以看出“
風速”和“溫度”重要性最
高，代表這兩個特徵輸入
值對於預測“用電量”結果
影響最大。

Importance	
windspeed	0.227763
temperature	0.201484
var1	0.147444
Month	0.131163
Week	0.083538
Day	0.052116
Year	0.047998
Time	0.032364
var2	0.026820
Weekday	0.025511
pressure	0.023800
Quarter	0.000000

特徵輸入值分佈狀況



用電量預測結果

測試集輸入

Xgboost演算法

ID	預測結果	實際用電量
0	208	216.0
1	226	210.0
2	229	225.0
3	212	216.0
4	186	222.0
5	244	216.0
6	236	219.0
7	237	222.0
8	214	219.0
9	229	222.0
10	203	201.0
11	208	204.0
12	219	210.0
13	223	216.0
14	220	216.0

MLP演算法

ID	預測結果	實際用電量
0	174	216.0
1	174	210.0
2	174	225.0
3	174	216.0
4	174	222.0
5	219	216.0
6	249	219.0
7	240	222.0
8	221	219.0
9	241	222.0
10	222	201.0
11	218	204.0
12	238	210.0
13	237	216.0
14	228	216.0